BootCamp Final Project

Amisi FIKIRINI AMISI
African Master in Machine Intelligence (AMMI), Senegal

Supervised by: Tutors

Project Report

April 9, 2023

African Institute for Mathematical Sciences
SENEGAL

Contents

1. Introduction

This report presents the results obtained after the experimentation within the framework of the final project of the BootCamp whose main objective was the learning and maturation of the notions required behind machine learning through theory and practice. The main objective of the project is to develop a supervised learning model solving a linear regression problem by using two approaches (normal equation and/or Classical Gram-Schmidt) to find the parameters allowing to make an approximate prediction of the reality of a given new observation, to compare the results of these approaches and to choose the best one.

1.1 The algorithms involved

As mentioned in the introduction, there are two approaches to compare the results, as follows:

1.1.1 Normal Equation

Normal Equation is an analytical approach to Linear Regression with a Least Square Cost Function. It is a closed-form solution used to find the value of the parameter θ that minimizes the cost function. Another way to describe the normal equation is as a one-step algorithm used to analytically find the coefficients that minimize the loss function. And it is given by :

$$\hat{y} = \theta_0 x_0 + \theta_1 x_1 + ... + \theta_n x_n$$

Where $\hat{y}$ is the hyphotesis function, θ represents the parameters (defined below) and n is the number of features.

$$\theta = (X^T X)^{-1} X^T Y$$

1.1.2 Classical Gram-Schmidt

The classical Gram-Schmidt process is a sequence of operations that allow us to transform a set of linearly independent vectors into a set of orthonormal vectors that span the same space spanned by the original set. The main goal of Gram-Schmidt is to transform a linearly independant vectors to orthonromal vectors.

$$S_1, ..., S_k = u_1, ..., u_k$$

That means, the finded orthonormal vectors representes our parameter θ, as it is also for theta in normal equation defined above.

2. Experimentation and Results

Here we are going to presente and make comparaison about the results that we obtain from those two approches.

2.1 Comparison by Mean Square Error (MSE)

Model Evaluation is a crucial aspect in the building of a model. When the purpose of the model is prediction, a reasonable parameter to validate the quality of model is the mean squared error of prediction.

MSE measures how close a regression line is to a set of data points. It is a risk function corresponding to the expected value of the squared error loss. It is calculated by taking the average, specifically the mean, of errors squared from data as it relates to a the hypothesis function.

2.1.1 MSE of Normal Equation approach

By following as well the approach process, the MSE found for this first approach (Normal Equation) is equal to **0.4978843746297357**. We can approximate to **0.498**.

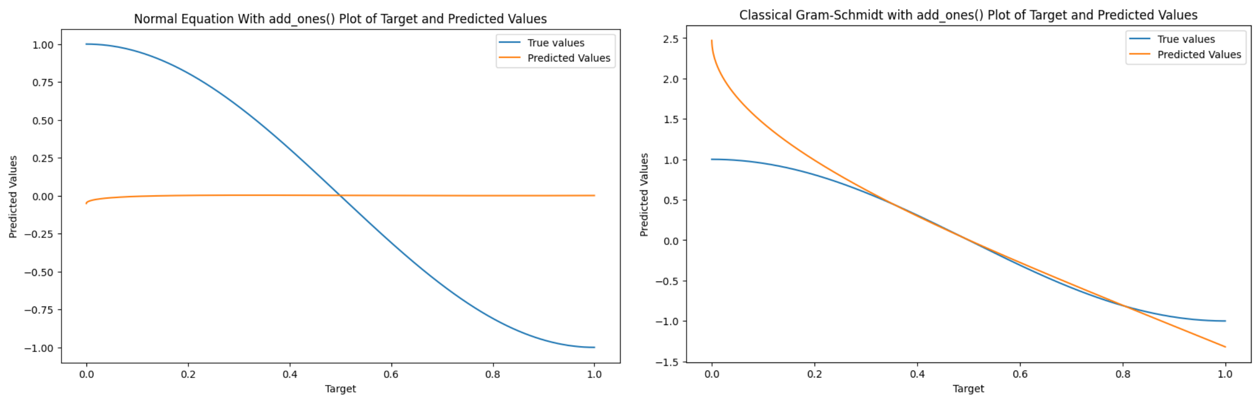
2.1.2 MSE of Gram-Schmidt

Here, the MSE found for this second approach (Gram-Schmidt) is equal to **1.3611326992212045** that we can approximate to **1.361**.

So, refere to the MSE computed for the both approaches, it is very clear that the Normal Equation is the best one because of the small mean of errors that can be occurs on the prediction.

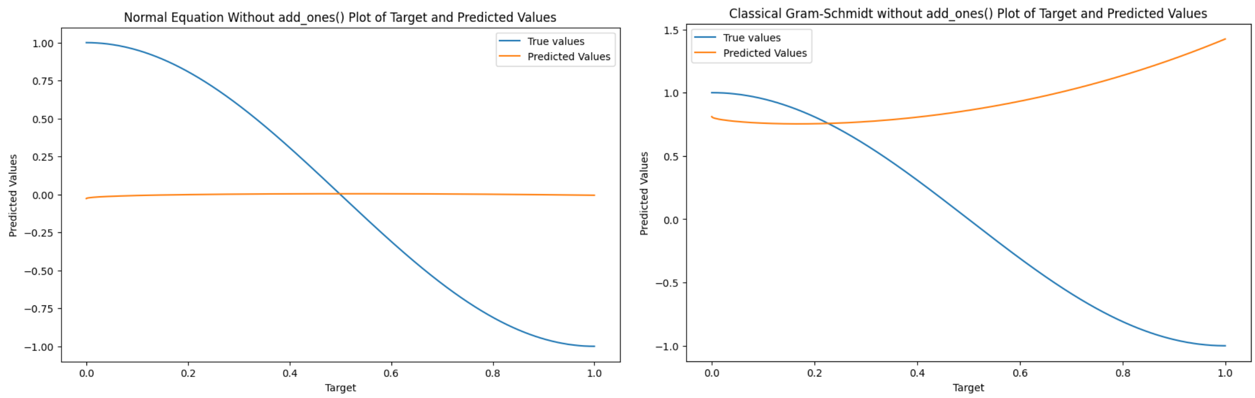
2.2 Comparison by Plot

As we can see, the results for Normal Equation stays the same while for Classical Gram-Schmidt are differents by applying or removing addOnes() methods. Thus, The Normal Equation stays our best approach.



(a) Normal Equation (b) Classical Gram-Schmidt

Figure 2.1: Normal Equation and Classical Gram-Schmidt with addOnes



(a) Normal Equation (b) Classical Gram-Schmidt

Figure 2.2: Normal Equation and Classical Gram-Schmidt without addOnes

2.3 The complexity of Gram-Schmidt.

For a given matrix(Dataset) of **m** rows and **n** columns, the complexity of the Gram-Schmidt orthogonalization algorithm is equal to : $2mn^2$ operations.

The overall complexity of Gram-Schmidt algorithm is $O(mn^2)$:

The process must be applied n times and each orthogonalization takes $O(mn)$ opérations (multiplications and additions) so altogether it makes $O(mn^2)$ complexity.

The QR factorization can be improved with both CGS (classical gram-schmidt) and MGS (modified gram-schmidt).

3. Conclusion

We are going to conclude that, the normal equation is the best one for our generated data.